OFFICE OF CIVILIAN RADIOACTIVE WASTE MANAGEMENT ANALYSIS/MODEL COVER SHEET

1. QA: QA

Page: 1 of: 34

Complete Only Applicable Items

2. ☒ Analysis — Check all that apply

Type of Analysis
- ☐ Engineering
- ☒ Performance Assessment
- ☐ Scientific

Intended Use of Analysis
- ☐ Input to Calculation
- ☒ Input to another Analysis or Model
- ☐ Input to Technical Document
- ☐ Input to Technical Products

Describe use: TSPA-SR Dose Calculation

3. ☐ Model — Check all that apply

Type of Model
- ☐ Conceptual Model
- ☐ Abstraction Model
- ☐ Mathematical Model
- ☐ System Model
- ☐ Process Model

Intended Use of Model
- ☐ Input to Calculation
- ☐ Input to another Model or Analysis
- ☐ Input to Technical Document
- ☐ Input to Technical Products

Describe use:

4. Title:
Distribution Fitting to the Stochastic BDCF Data

5. Document Identifier (including Rev. No. and Change No., if applicable):
ANL-NBS-MD-000008, Rev. 00

6. Total Attachments: 3

7. Attachment Numbers - No. of Pages in Each:
I-1, II-13, III -CD

	Printed Name	Signature	Date
8. Originator	Anthony J. Smith	AJ Smith	14APR00
9. Checker	Noreen Baker	WR Hunt for N. Baker	4/14/00
10. Lead/Supervisor	Anthony J. Smith	AJ Smith	14 APR 0
11. Responsible Manager	Jerry A. McNeish	Jerry A McNeish	4·14·00

12. Remarks:
A CD is included with this document as Attachment III. The "Excel Files" directory contains the Microsoft ® Excel 97 SR-2 files discussed in the document.

INFORMATION COPY
LAS VEGAS DOCUMENT CONTROL

NM5507

Enclosure 9

OFFICE OF CIVILIAN RADIOACTIVE WASTE MANAGEMENT
ANALYSIS/MODEL REVISION RECORD
Complete Only Applicable Items

1. Page: 2 of: 34

2. Analysis or Model Title:
Distribution Fitting to the Stochastic BDCF Data

3. Document Identifier (including Rev. No. and Change No., if applicable):
ANL-NBS-MD-000008, Rev. 00

4. Revision/Change No.	5. Description of Revision/Change
Revision 00	Initial Issue

CONTENTS

FIGURES

TABLES

1. PURPOSE

The objective of this analysis is to derive statistical approximations (abstractions) to the individual data-sets of the Biosphere Dose Conversion Factors (BDCFs). These abstractions will be used in the Total System Performance Assessment (TSPA) code for the proposed Yucca Mountain repository and, if necessary, in further analysis of the soil build-up phenomenon prior to incorporation into the TSPA calculations. Each data set comprises of 130 stochastic realizations of BDCFs evaluated for a given radionuclide after a predefined period of previous irrigation. Each individual realization is generated by sampling the input parameters over their region of uncertainty. The detail of this BDCF generation is the subject of the report titled *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000).

Several statistical distributions will be evaluated against a small subset of the total set of BDCFs to give some indication of those that could be considered suitable and eliminate any that are statistically unacceptable. However, it should be emphasized that the goal is not necessarily to identify the optimum distribution for each BDCF set. The objective is to identify a single distribution that provides an acceptable statistical fit to all sets of BDCFs for a given radionuclide. The desire to identify a single distribution for each radionuclide arises when soil build up causes a significant change over time. A significant change is one for which a conservative approach, that can be advocated when the build-up is small, could be unacceptable in cases where there are large changes of BDCFs over time. If one distribution can be identified for each radionuclide, the fitting of a time evolution function to parametric data becomes a more simple goal that can be more readily justified. Soil build-up occurs when the radionuclide concentration in soil increases with the period of irrigation. The dose from pathways that transport radionuclides from the soil to the receptor continues to increase until equilibrium is established in the soil. Whether or not soil build-up is a significant effect for any radionuclide is dependent upon details of the biosphere model being used and the habits of the critical group. The details of the scenario used to generate the BDCFs are presented in the Analysis and Modeling Report (AMR) *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000). In cases where build-up is a significant process, this abstraction of the previous period of irrigation allows for consideration of atmospheric soil erosion in a subsequent AMR. (Erosion is a process not considered by the methodology used in CRWMS M&O 2000).

The activities described in this report were conducted in accordance with the Work Direction and Planning Document titled *Abstraction of BDCF Distributions* (CRWMS M&O 1999a).

2. QUALITY ASSURANCE

The quality assurance (QA) program applies to the development of this analysis documentation. The information provided in this analysis will be used for evaluating the post-closure performance of the Monitored Geologic Repository (MGR) waste package and engineered barrier segment. The Performance Assessment Operations (PAO) responsible manager has evaluated the technical document development activity in accordance with QAP-2-0, *Conduct of Activities*. The QAP-2-0 activity evaluation (CRWMS M&O 1999b) has determined that the preparation and review of this technical document is subject to *Quality Assurance Requirements and Description* (DOE 2000) requirements. The activity evaluation (CRWMS M&O 1999b) remains in effect even though QAP-2-0, *Conduct of Activities*, has been superseded by

AP-2.16Q, *Activity Evaluation*. The effort reported in this AMR was conducted and documented in accordance with AP-3.10Q, *Analyses and Models* and AP-3.15Q, *Managing Technical Product Inputs*. A work plan was developed, issued, and utilized in the preparation of this document (CRWMS M&O 1999a). Since the analysis does not involve any field activity, there is no determination of importance evaluation developed in accordance with NLP-2-0, *Determination of Importance Evaluations*. There are no permanent items addressed in this AMR, so it is not subject to QAP-2-3, *Classification of Permanent Items*.

An evaluation of the control of electronic management of data has been performed for this activity (per AP-SV.1Q), and it was concluded that current processes are adequate to ensure the accuracy, completeness, and security of the data used in this activity.

3. COMPUTER SOFTWARE AND MODEL USAGE

No models were used or developed in this analysis. The only software used was a commercially available spreadsheet (Microsoft® Excel 97 SR-2). This spreadsheet was used as an aid in calculation; no routines, macros, or other applications were developed and used. Use of this software is documented in Attachment II in accordance with AP-SI.1Q, *Software Management*.

4. INPUTS

4.1 DATA AND PARAMETERS

The data used in this analysis are reported in *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000). The data were taken from the Technical Database Management System (TDMS) with Data Tracking Number (DTN) MO0004SPABDCFS.001. It should be noted that each of the six files (with extensions *.flg, *.inp, *.out, *.pti, *.rst, and *.vec) associated with one calculation of the stochastic "Realistic Representation" (file name starts with Rr) represents either the input or output files for the GENII-S code (Leigh et al. 1993). A set of the input and output files exist for each calculation comprising of six period of previous irrigation (to address radionuclide build-up in soils) for 18 radionuclides. The results output file (*.rst) contain the requested 130 individual stochastic realizations for each set of BDCFs. Each realization is generated by randomly sampling from the distributions representing the uncertainty in the numerical values for those parameters where this uncertainty has been defined in CRWMS M&O 2000. In addition to uncertainties in results due to parametric uncertainties, the reader should be made aware that as in all modeling effort there are other sources of uncertainty. These other sources include:

- Uncertainty in the conceptual model developed to represent the actual situation being modeled.
- Uncertainty in the mathematical implementation of conceptual model.
- Numerical uncertainty in the computer solution to the mathematical model.
- Uncertainty in applying the computer model.

By opting to use the GENII-S code, it is implicit that the uncertainties and errors from these and other sources are of no consequence (when compared to the uncertainties arising from the parametric uncertainties). This assumption will have to be assessed by the code validation process discussed in Section 3 of CRWMS M&O 2000.

The output file contains multiple columns of data showing the values of stochastically sampled input parameters and the multiple attendant predicted doses (such as the dose to each organ and the external dose, effective dose equivalent and total dose). Only the last column of the data with heading TEDE (Total Effective Dose Equivalent) presents the raw BDCF data that is processed in this AMR. The input file with extension *.inp provides the number of years of prior irrigation used for that specific calculation.

4.2 CRITERIA

This AMR was prepared to conform with DOE interim guidance (Dyer 1999) which directs the use specified Subpart/Sections of the proposed NRC high-level waste rule, 10 CFR Part 63 (64 FR 8640). Specified Subparts of this proposed rule that are particularly applicable to data include Subpart B, Section 15 (Site Characterization) and Subpart E, Section 114 (Requirements for Performance Assessment).

The U.S. Nuclear Regulatory Commission's (NRC's) Total System Performance Assessment and Integration (TSPA&I) Issue Resolution Status Report (IRSR) (NRC 1998) establishes generic technical acceptance criteria. These criteria are considered by the NRC staff to be essential to a defensible, transparent, and comprehensive assessment methodology for the repository system. These regulatory acceptance criteria address five fundamental elements of the Department of Energy (DOE) TSPA model for the Yucca Mountain site, namely:

1. Data and model justification (focusing on sufficiency of data to support the conceptual basis of the process model and abstractions)
2. Data uncertainty and verification (focusing on technical basis for bounding assumptions and statistical representations of uncertainties and parameter variability)
3. Model uncertainty (focusing on alternative conceptual models consistent with available site data)
4. Model verification (focusing on testing of model abstractions using detailed process-level models and empirical observations)
5. Integration (focusing on appropriate and consistent coupling of model abstractions).

Relevant to the topic of this AMR, elements (1) through (4) of the acceptance criteria were used to generate the individual sets of stochastic BDCF data as reported in CRWMS M&O 2000. This AMR reduces the large volume of data described in 4.1 into a simplified statistical form for use in the TSPA-SR predictive capability. The process must preserve the integrity of the data (elements 1 & 4) while retaining the uncertainty inherent in the biosphere model/data (elements 2 & 3). The process conducted by this AMR is part of element (5) of the NRC acceptance criteria.

4.3 CODES AND STANDARDS

There are no applicable codes or standards.

5. ASSUMPTIONS

5.1 GOODNESS OF FIT

It is assumed that the elementary statistical test known as the *Chi Square* (sometimes referred to as *Chi Squared*) test is adequate to demonstrate acceptable distribution fits to the stochastic BDCF data. It is acknowledged that alternative statistical tests are available, however for this task the standard Chi Square test for testing goodness of fit was used.

5.2 SOIL BUILD-UP

The BDCFs discussed in 4.1 are provided as a function of previous irrigation periods. The more contaminants that are added to agricultural land, by virtue of continuing irrigation using unit concentration of radionuclides, the greater becomes the expected BDCF value. The actual magnitude (and therefore significance) of this build-up is dependent on the inputs used as reported in *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000). For the data defined in 4.1, the magnitude of build-up spans the range from less than one percent (for ^{129}I) up to over 200 percent (for ^{229}Th). For the purpose of the analysis reported here, soil build-up effects were only considered in detail if the magnitude was greater than 15 percent. For those radionuclides for which build-up was not considered, the BDCF abstractions were generated for the longest period of previous irrigation. Thus, these BDCFs were upper bounding values and therefore conservative. Being conservative (i.e., leading to dose overestimates and not underestimates), this assumption is considered reasonable for use.

6. ANALYSIS/MODEL

The purpose of this AMR is to determine an acceptable statistical distribution (or distributions) and the defining parameters to represent the empirical distributions of the BDCFs generated by the effort reported in the AMR titled *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000). As discussed in Section 4.1, the empirical distributions (CRWMS M&O 2000) reflect the resultant uncertainties in the BDCFs arising from to the uncertainties in some of the GENII-S input parameters. The details of these uncertainties and the parameters to which they apply are presented in CRWMS M&O 2000. The details of the approach used to capture the uncertainty reflected in the empirical BDCFs generated by CRWMS M&O 2000 in an abstracted statistical distribution are presented in the following sections.

6.1 OVERVIEW OF THE APPROACH

6.1.1 Statistical Distributions Considered

There are numerous statistical distributions that could be tested to determine whether they provide a sufficiently good fit to the empirical data. While this may have been an interesting exercise, such a rigorous approach was not necessary, as it was only required to determine whether a distribution provided an acceptable fit in the statistical sense. For this reason, it was elected to initially only consider a limited set of distributions to determine whether ready acceptance could be demonstrated. If acceptable distribution could not be determined for each BDCF data set, then it was accepted that additional distributions would have to be considered. The initial set of statistical distributions considered for approximating the BDCF data were the:

- Normal
- Lognormal
- Triangular
- Weibull
- Shifted lognormal distribution.

The normal, lognormal, and Weibull distributions each require two parameters to characterize them. All three distributions are available as (statistical) functions within the Excel spreadsheet software (Microsoft 1997 – Start the Excel spreadsheet program, click on the *Help* pull down menu, select *Contents and Index,* select *Index* Tab, enter function name, press *Display* button). The triangular and shifted lognormal are not available as explicit functions within Excel and require a little more explanation (see 6.1.1.1 & 6.1.1.2).

For the lognormal distribution (and by a simple axis translation, the shifted lognormal distribution), the parameter definition used in this report was that implemented in the Excel Spreadsheet. If x is distributed with a lognormal distribution, then the mean of the distribution is defined to be the mean value of $ln(x)$. In addition, the standard deviation of the distribution is the standard deviation of $ln(x)$.

6.1.1.1 Triangular Distribution

The probability density function for a triangular distribution is a triangle. In general, three parameters (a, b, and c) are needed to characterize such a distribution. These three parameters are lower and upper limit of the distribution and the mode (i.e., peak) of the distribution. Figure 1 shows a graphical representation of such a probability distribution function (pdf). For values of the random variable less than a and greater than c, the probability of the event is zero. The mode of the distribution is at b.

As this is a probability distribution, the area (integral) under the probability distribution curve has to be unity. This fact allows h to be defined in terms of the other parameters. Elementary geometrical considerations (the area of a triangle is given by half the product of the base and the perpendicular height) lead to the following expression relating parameters:

$$\frac{h(c-a)}{2}=1$$

(Eq. 1)

or

$$h=\frac{2}{(c-a)}$$

(Eq. 2)

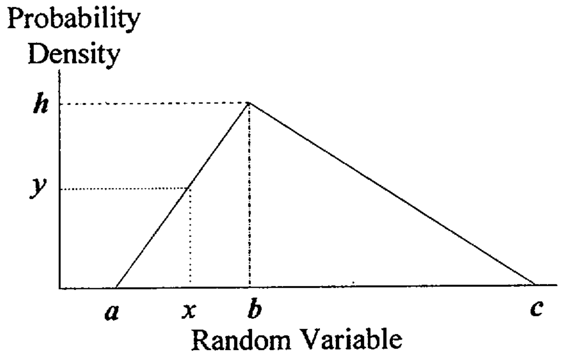

Figure 1. Defining Parameters of the Triangular Distribution

For this analysis, the cumulative distribution function (cdf) is required. For a random variable x that is greater than a but less than b, the cdf is, from elementary geometry considerations of the area of a triangle, half the base $(x\text{-}a)$ times the height y. Simple scaling of similar triangles gives,

$$y=\frac{(x-a)h}{(b-a)}$$

(Eq. 3)

Making the substitution for y provides the cdf,

$$cdf=\frac{(x-a)^2h}{(b-a)}$$

(Eq. 4)

Substituting for h as determined above in Equation 2 gives,

$$cdf = \frac{(x-a)^2}{(c-a)(b-a)}$$

(Eq. 5)

When the variable x is between b and c, the same approach gives the following equation,

$$cdf = 1 - \frac{(c-x)^2}{(c-a)(c-b)}$$

(Eq. 6)

6.1.1.2 Shifted Lognormal Distribution

The lognormal distribution occurs when the natural logarithm (i.e., to base e) of the random variable is distributed normally, i.e., $ln(x_i)$ is normal with a defined mean (μ) and standard deviation (σ). A shifted lognormal distribution occurs when $ln(x_i\text{-}s)$ (where s is a constant) is distributed normally with a defined mean (μ) and standard deviation (σ). Thus for a shifted lognormal distribution, the three parameters (s, μ, and σ) required to characterize the distribution are the shift (or axis translation) s, the mean, μ, of $ln(x_i\text{-}s)$, and the standard deviation, σ, of $ln(x_i\text{-}s)$.

6.1.2 Statistical Fitting

The χ^2 test for goodness of fit was used to determine whether a distribution provided an acceptable fit to the BDCF data. The details of this test were taken from a standard text on elementary statistics (Bulmer 1979, p. 154 - 161). The salient details of the method are provided below.

Let there be n realizations (observations) that can be grouped into k classes, such that the i^{th} class contains n_i observations. If E_i are the expected number of observations predicted by the distribution under test, then the χ^2 criterion of goodness of fit is defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$$

(Eq. 7)

If the distribution under test completely predicts the observed data, then for all i, $E_i = n_i$ and $\chi^2 = 0$. Such an exact fit would be unlikely. Even if the observations were sampled from the postulated parent distribution, it would be expected that there would be differences between the expected (real numbers) and observed (integer). However, as the predictive capability of the distribution under test becomes less good, the difference between the observed and expected numbers become increasingly larger and the larger χ^2 becomes. The expected numerical values of χ^2 are further discussed and quantified in the next section.

Bulmer (1979, p. 156) proceeds to show that the χ^2 criterion approximately follows the χ^2 distribution with k-1-p degrees of freedom, where p is the number of parameters which have been independently estimated from the data. The parameters that have to be estimated are those required to define the distribution under consideration. For the approximation to hold, it is necessary that the number of predicted observations in any interval should not be too small. Bulmer (1979, p. 158) advises that it has been found empirically that the approximation is satisfactory provided each selected interval is predicted to have five or more observations.

6.1.3 Significance Test

To derive a significance test, the approach of Bulmer (1979, p. 160) is followed. The hypothesis is made that the postulated distribution is representative of the data (observations). As small values of χ^2 indicate good agreement with the hypothesis, the hypothesis should only be rejected when χ^2 is large. The χ^2 distribution is used to generate the probabilities of obtaining a value of χ^2 greater than the observed value on the assumption that the hypothesis is true. If repeated sampling were to be performed, the χ^2 criterion would follow the χ^2 distribution with the appropriate number of degrees of freedom. If this probability is small, the hypothesis is rejected, otherwise it is accepted.

Table 1 gives the values of χ^2 distribution which are exceeded with probability P for a range of probabilities and degrees of freedom. This table was generated in Excel using the "CHIINV" statistical function that provides (Microsoft 1997) *the inverse of the one-tailed probability of the chi-squared distribution* and copied into this document. The values of χ^2 presented for the values of P common to both Bulmer (1979, p. 234) and Table 1 (i.e., values of P of 0.05 and 0.01) are identical.

Table 1. Percentage Points of the χ^2 Distribution

Degrees of Freedom	**Value of *P***									
	0.50	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.002**
6	5.3	6.2	7.2	8.6	10.6	12.6	14.4	16.8	18.5	20.8
7	6.3	7.3	8.4	9.8	12.0	14.1	16.0	18.5	20.3	22.6
8	7.3	8.4	9.5	11.0	13.4	15.5	17.5	20.1	22.0	24.4
9	8.3	9.4	10.7	12.2	14.7	16.9	19.0	21.7	23.6	26.1
10	9.3	10.5	11.8	13.4	16.0	18.3	20.5	23.2	25.2	27.7
11	10.3	11.5	12.9	14.6	17.3	19.7	21.9	24.7	26.8	29.4
12	11.3	12.6	14.0	15.8	18.5	21.0	23.3	26.2	28.3	31.0
13	12.3	13.6	15.1	17.0	19.8	22.4	24.7	27.7	29.8	32.5
14	13.3	14.7	16.2	18.2	21.1	23.7	26.1	29.1	31.3	34.1
15	14.3	15.7	17.3	19.3	22.3	25.0	27.5	30.6	32.8	35.6

Say an experiment was performed, a distribution for the data was postulated, and the resulting value of χ^2 was 10.0 for 10 degree of freedom. Then the values of the χ^2 distribution given in Table 1 indicate that it would be expected that repeating the experiment would result in approximately 45% of the values of χ^2 would be greater than 10. Thus, the proposed distribution would be accepted. If the experiment had yielded a value of χ^2 of 26, then Table 1 tells us that if the experiment were repeated multiple times only 1 test in a few hundred would result in such a high value. Thus the hypothesis would be rejected (i.e., the proposed distribution does not represent the data) at the 1% level. That is there is little chance that the observed high value of χ^2 is due to the random nature of the test. It can therefore be taken that the distribution does not adequately represent the parent distribution from which the data were sampled. As will be seen, about 50 distributions were evaluated in this AMR. At the 10% level, it would be anticipated that approximately 5 of the 50 would result in values of χ^2greater than the 10% value. This is the case.

It should be mentioned that other tests exist to determine the adequacy of an approximating distribution to predict a set of stochastic data. These techniques may be used in later revisions of this effort. For this analysis the standard statistical χ^2 test is used.

6.1.4 Structure of Test

As mentioned in 4.1, there are 130 individual BDCF data points in each output file. (There is one file for each radionuclide at every defined irrigation time.) In 6.1.2, it was stated that to apply the χ^2 test, these data have to be grouped into a number of categories (i.e., bins of BDCFs between defined limits). In addition, the number and size (width) of these categories has to be selected such that, for the approximating distribution there are five or more predicted observation in all intervals.

For a continuous random variable, such as the BDCFs under consideration, there are an infinite number of ways of defining the bin structure. However, two simple approaches are available. The first uses a number of bins of equal size (width), where the number and size of the bins are selected such that there are more than the minimum number of observations (>5) in each bin. For BDCF that are by necessity positive, the lower and upper bins are bounded by 0 (other distributions could extend to $-\infty$) and $+\infty$ respectively.

An alternative approach is to divide the random variable space into the required number of intervals each containing the same number of observations. In the case of the BDCFs inputs to this effort where there are 130 data points, the logical choice would be to have 10 intervals containing 13 points or 13 intervals containing 10 points. Other data sets with different number of points can easily be accommodated by allowing one or more bins (easily done for the lowest and highest bins) to contain a different number of observations than the other bins.

The available functions within the Excel spreadsheet made the latter approach the easiest to implement. This was used in this analysis. Bin boundaries were defined such that each of 10 bins contained 13 stochastic BDCFs.

For each of the distributions considered (see below), input parameters were estimated that provided an approximate fit to the data. The $\frac{(observed - expected)^2}{expected}$ variable was calculated for each bin considered. The sum of these values provided the χ^2 value that would be used to determine the adequacy of fit. The built in Excel optimization function SOLVER (in pull down menu TOOLS) was used to determine the input parameters that resulted in a minimum value for the χ^2 value. SOLVER is an integral part of the Excel spreadsheet. The user sets up a spreadsheet that takes values in defined cells (inputs) to calculate an output value. Then SOLVER provides the user with the capability to determine the numerical values of the input cells that correspond to extreme values (maximum or minimum as selected by the user) of the output cell. When using SOLVER it was observed in some cases where the initial estimates for the inputs gave a high χ^2 value, the SOLVER function could not converge while simultaneously varying multiple inputs. In these cases, SOLVER had to be used to optimize individual parameters one at a time until a reasonable χ^2 value was achieved. Once the multiple single dimension (one parameter at a time) iterative use of SOLVER achieved an approximate fit, SOLVER could then be used to simultaneously optimize all parameters.

6.2 DETAILS OF THE ANALYSIS

The purpose of this analysis was to provide the TSPA code with an abstraction of the BDCF data. The abstraction will allow the sampling of the predicted distribution for the 18 radionuclides evaluated in *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000). The abstractions are to be undertaken over a period of time for which build-up effects are important. What was not known during the planning phase, was whether soil build-up would be a significant effect or not for any or all of the radionuclides considered for the defined receptor. In addition, only the site and scenario specific BDCF data could be used to determine which of the approximating distributions could be considered adequate in the statistical sense. To avoid unnecessary effort in completely analyzing each and every of the BDCF data sets, the data were first reviewed to determine whether soil build-up effects were significant. This review considered the ratio of the mean BDCF at the longest period of previous irrigation to the mean BDCF for no previous irrigation. For TSPA-SR, the conservative approach of using the BDCFs generated for the longest irrigation period was taken for those radionuclides where soil build-up effect were less than 15%. (Table 4 of CRWMS M&O 2000 shows the standard deviation of the BDCF distributions was approximately 15% to 20% of the mean value of the distribution, so the sampled BDCFs could be expected to vary over a range of about ±40%.) Second, a subset of BDCF data were used to perform the χ^2 statistical goodness of fit test for each of the distributions considered. This allowed any unsuitable distribution to be eliminated from further consideration and for attention to be focussed on the better fitting distributions.

The BDCF data were made available in two parts. The first data set contained the information on ^{227}Ac, ^{241}Am, ^{243}Am, ^{14}C, ^{129}I, ^{237}Np, ^{238}Pu, ^{239}Pu, ^{240}Pu, ^{99}Tc, ^{229}Th, ^{232}U, ^{233}U, ^{234}U, ^{236}U, and ^{238}U. These radionuclides were used to generate the report given below with placeholders for the two radionuclides (^{137}Cs and ^{90}Sr), the data for which, for logistic reasons, were supplied some two months after the initial data. Thus, the data for ^{137}Cs and ^{90}Sr were not available for consideration during the initial distribution analysis and inclusion in Table 6. It should be noted (see Section 6.2.1.5) that Table 6 show the results of the initial screening of acceptable distributions that was conducted on a subset of radionuclides for which data existed at that time.

6.2.1 Input Data

The GENII-S input/output data files are available under DTN MO0004SPABDCFS.001. The file naming convention for the data is as follows. The top-level directory is named "Bdcf." Within this directory there are two secondary directories "Rr_data" and "Sc_data." The initial two letters of these directories stand for "reasonable representation" and "safety case" respectively. The former are based on best estimates of the parameters used as input to the stochastic GENII-S, while the latter contains deterministic BDCFs based on upper limit (and thereby very conservative) input parameter values. TSPA at this stage is only concerned with the most reasonable data available i.e., those within the "Rr_data" directory.

The data files, input and output, are contained with this "Rr_data" directory. The naming convention for the files containing the data for ^{227}Ac, ^{241}Am, ^{243}Am, ^{14}C, ^{129}I, ^{237}Np, ^{238}Pu, ^{239}Pu, ^{240}Pu, ^{99}Tc, ^{229}Th, ^{232}U, ^{233}U, ^{234}U, ^{236}U, and ^{238}U is as follows:

- The first two characters are "Rr", again denoting reasonable representation.
- The next character is an integer and has a value of 1 to 6. This is the integer representing the run for the previous period of irrigation for that given radionuclide. "1" always represents "no previous periods of irrigation with contaminated water." Table 2 gives the periods of previous irrigation used for each of the calculations.
- The next one or two alphabetic character(s) is the accepted elemental chemical symbol for the radionuclide under consideration. Only carbon (C), iodine (I), and uranium (U) have a single character.
- The final block of (two or three numeric) characters are the values of the atomic weight of the radionuclide under consideration.
- For each radionuclide, there are six files (same file name) with different extensions. These extensions are *.flg, *.inp, *.out, *.pti, *.rst, and *.vec.

For ^{137}Cs and ^{90}Sr the naming convention was different (being done at a later time by a different person).

- The first two characters are "1b".
- The final set of characters is either "cs137" or "sr90" (this is self-explanatory).
- If there are no other characters then the data apply to irrigation period 4.
- If there is an additional character (1, 2, 3, 5, or 6), then the data in the file are appropriate to that period of irrigation period.

6.2.1.1 Periods of Prior Irrigation

For each radionuclide, a set of six periods of previous irrigation that should be considered were defined based upon the estimated soil leaching factor for that element and the half-life of the radionuclide under consideration.

The actual values used in the GENII-S calculations by the AMR *Non-Disruptive Event Biosphere Dose Conversion Factors* (CRWMS M&O 2000) are contained in the "*.inp" file. This file contains many user defined input data for the code. The contents of the file can be viewed by opening with one of the many available word processing programs. The 42nd line of this file contains the text *"H2O Rel Time Before Intake yr"*. The following line contains a string of "-". The next line (44th) line contains the number of years of previous irrigation used for the calculation associated with the file name. It should be emphasized that the work conducted in the AMR made use of data generated elsewhere by the use of the GENII-S code. The GENII-S code was not used in any analyses reported in this AMR.

The periods of previous irrigation used in generating the BDCFs are given in Table 2.

Table 2. Number of Years of Previous Irrigation for each Radionuclide

Radionuclide	Period of Previous Irrigation					
	1	2	3	4	5	6
	(Values in Table are Years)					
^{227}Ac	0	6	13	22	35	56
^{241}Am	0	114	253	432	685	1117
^{243}Am	0	511	1138	1947	3084	5031
^{14}C	0	752	1674	2864	4537	7401
^{137}Cs	0	8	18	30	48	78
^{129}I	0	1	2	3	4	5
^{237}Np	0	1	3	5	8	14
^{238}Pu	0	23	51	88	139	227
^{239}Pu	0	148	329	563	893	1456
^{240}Pu	0	148	329	563	893	1456
^{90}Sr	0	5	12	21	33	53
^{99}Tc	0	1	2	3	4	5
^{229}Th	0	858	1910	3269	5179	8448
^{232}U	0	9	21	36	57	93
^{233}U	0	9	21	36	57	93
^{234}U	0	9	21	36	57	93
^{236}U	0	9	21	36	57	93
^{238}U	0	9	21	36	57	93

DTN: MO0004SPABDCFS.001

6.2.1.2 Stochastic BDCF Data

The output from the GENII-S code for all data for each of the 130 realizations is contained in the "*.rst" file. The origins of these data are discussed in Section 4.1 where the originating AMR is discussed and the DTN for the data is provided DTN:MO0004SPABDCFS.001). This file contains headers with file identification information followed by multiple blocks of five columns of data. The BDCF data are in the last column and have a header of *Annual EDE.* Because of the volume of data for just the BDCFs, these data are not presented in this AMR. They are contained in the Excel data files accompanying this report.

The BDCF data used in this AMR were imported into column B of the Excel spreadsheets. The external dose components were also imported (column A) although these data were not used in the analysis reported in this AMR. Each radionuclide was allocated a file with designator starting with the elemental chemical symbol (one or two characters) followed by the atomic mass of the radionuclide (two or three numerical characters) e.g., Am241, C14, and I129. These files are provided on the attached media. The names have been modified in these attachments by adding the data of the last working change (i.e., analysis) and the work "final" to show that all the intended analysis had been completed. Within a given file, each period of previous irrigation was allocated a worksheet with the same designator as defined in 6.2.1.

6.2.1.3 Mean and Standard Deviation of the BDCF Data

An additional worksheet was inserted into each of the workbook files defined in 6.2.1.2. The raw BDCF data from each of the other six sheets of the workbook were copied by reference into adjacent columns starting at row 9. The mean and standard deviation for each column of were generated by using the Excel functions "AVERAGE" and "STDEV." This information is shown in Tables 3 and 4.

Table 3. Mean of 130 Realizations of BDCFs

	BDCF Mean (rem/year per picoCurie/liter)					
	Previous Irrigation Period					
Radionuclide	**1**	**2**	**3**	**4**	**5**	**6**
^{227}Ac	1.81E-02	1.81E-02	1.81E-02	1.82E-02	1.82E-02	1.83E-02
^{241}Am	4.65E-03	4.74E-03	4.83E-03	4.91E-03	4.99E-03	5.05E-03
^{243}Am	4.64E-03	5.57E-03	6.47E-03	7.30E-03	8.07E-03	8.74E-03
^{14}C	4.05E-06	4.05E-06	4.05E-06	4.05E-06	4.05E-06	4.05E-06
^{137}Cs	8.77E-05	1.09E-04	1.31E-04	1.52E-04	1.73E-04	1.94E-04
^{129}I	3.61E-04	3.61E-04	3.62E-04	3.62E-04	3.62E-04	3.62E-04
^{237}Np	6.76E-03	6.77E-03	6.78E-03	6.79E-03	6.80E-03	6.81E-03
^{238}Pu	4.11E-03	4.12E-03	4.13E-03	4.14E-03	4.16E-03	4.17E-03
^{239}Pu	4.56E-03	4.65E-03	4.75E-03	4.85E-03	4.94E-03	5.03E-03
^{240}Pu	4.55E-03	4.65E-03	4.74E-03	4.83E-03	4.92E-03	5.01E-03
^{90}Sr	1.82E-04	2.26E-04	2.71E-04	3.06E-04	3.33E-04	3.51E-04
^{99}Tc	4.02E-06	4.08E-06	4.08E-06	4.08E-06	4.08E-06	4.08E-06
^{229}Th	4.58E-03	7.55E-03	1.03E-02	1.27E-02	1.48E-02	1.65E-02
^{232}U	1.71E-03	1.79E-03	1.90E-03	1.99E-03	2.07E-03	2.13E-03
^{233}U	3.77E-04	3.78E-04	3.81E-04	3.83E-04	3.85E-04	3.88E-04
^{234}U	3.70E-04	3.72E-04	3.74E-04	3.76E-04	3.78E-04	3.80E-04
^{236}U	3.50E-04	3.52E-04	3.54E-04	3.56E-04	3.58E-04	3.60E-04
^{238}U	3.38E-04	3.41E-04	3.45E-04	3.48E-04	3.52E-04	3.55E-04

Table 4. Standard Deviation of 130 Realizations of BDCFs

	BDCF Standard Deviation (rem/year per picoCurie/liter)					
	Previous Irrigation Period					
Radionuclide	**1**	**2**	**3**	**4**	**5**	**6**
^{227}Ac	2.87E-03	2.87E-03	2.87E-03	2.88E-03	2.88E-03	2.88E-03
^{241}Am	7.38E-04	7.39E-04	7.42E-04	7.47E-04	7.50E-04	7.53E-04
^{243}Am	7.36E-04	8.14E-04	9.88E-04	1.20E-03	1.41E-03	1.62E-03
^{14}C	2.43E-07	2.43E-07	2.43E-07	2.43E-07	2.43E-07	2.43E-07
^{137}Cs	2.37E-05	2.73E-05	3.36E-05	4.06E-05	4.88E-05	5.71E-05
^{129}I	6.53E-05	6.54E-05	6.54E-05	6.55E-05	6.55E-05	6.55E-05
^{237}Np	1.07E-03	1.07E-03	1.07E-03	1.07E-03	1.07E-03	1.08E-03
^{238}Pu	6.52E-04	6.52E-04	6.52E-04	6.52E-04	6.52E-04	6.52E-04
^{239}Pu	7.24E-04	7.26E-04	7.28E-04	7.30E-04	7.33E-04	7.37E-04
^{240}Pu	7.23E-04	7.25E-04	7.27E-04	7.28E-04	7.31E-04	7.34E-04
^{90}Sr	3.58E-05	8.83E-05	1.50E-04	2.02E-04	2.41E-04	2.68E-04
^{99}Tc	1.63E-06	1.71E-06	1.72E-06	1.72E-06	1.72E-06	1.72E-06
^{229}Th	7.44E-04	1.40E-03	2.37E-03	3.27E-03	4.07E-03	4.74E-03
^{232}U	2.72E-04	2.77E-04	2.89E-04	3.07E-04	3.25E-04	3.41E-04
^{233}U	5.98E-05	6.00E-05	6.00E-05	6.02E-05	6.05E-05	6.08E-05
^{234}U	5.88E-05	5.89E-05	5.89E-05	5.91E-05	5.93E-05	5.96E-05
^{236}U	5.57E-05	5.58E-05	5.58E-05	5.61E-05	5.62E-05	5.65E-05
^{238}U	5.43E-05	5.45E-05	5.45E-05	5.48E-05	5.50E-05	5.53E-05

6.2.1.4 Significance of Radionuclide Build-up in Soils

To determine whether radionuclide accumulation effects in the soil are important for the scenario used to generate the stochastic BDCFs, an elementary calculation was performed. The ratio of the mean BDCF after the final defined period of irrigation to that for the first period (no previous irrigation) was calculated. This parameter provides a measure of the degree of the radionuclide build-up effect in the soil as predicted by the AMR providing the stochastic BDCF data (CRWMS M&O 2000). The resulting data was then sorted (using the Excel "SORT" function) to give the radionuclides in descending order of this (importance) ratio. The results of these operations are provided in Table 5.

By inspection of the data presented in Table 3 and 4, it can be seen that the standard deviation of the BDCF distributions is approximately 15%. This provides a measure of the uncertainty of the BDCFs due to parametric variability of 1.15. In Table 5, many of the radionuclides have build-up factors that are less (or significantly less) than the expected variability. Thus for these radionuclides (i.e., those up through ^{239}Pu in Table 5), the effect of soil build-up was small. For these radionuclides, the conservative approach of using the BDCFs appropriate to longest period of previous irrigation was adopted.

For the remaining five radionuclide, the effect of soil build-up was significantly greater and of potential concern. These radionuclides will be considered in a later analysis, where the potentially mitigating effect of soil erosion will be taken into account.

Table 5. Build-up Factors for Radionuclides

Radionuclide	Ratio of Periods 6 to 1	Period 6 Time (Years)
^{229}Th	3.60	8448
^{137}Cs	2.21	78
^{90}Sr	1.93	53
^{243}Am	1.88	5031
^{232}U	1.25	93
^{239}Pu	1.10	1456
^{240}Pu	1.10	1456
^{241}Am	1.09	1117
^{238}U	1.05	93
^{233}U	1.03	93
^{234}U	1.03	93
^{236}U	1.03	93
^{99}Tc	1.02	5
^{238}Pu	1.01	227
^{227}Ac	1.01	56
^{237}Np	1.01	14
^{129}I	1.00	5
^{14}C	1.00	7401

6.2.1.5 Distribution Fitting

As mentioned in 6.1.1, the statistical distributions considered for approximating the BDCF data were the normal, the lognormal, the triangular, the Weibull, and the shifted lognormal distributions. Initial calculations were performed on a select number of radionuclides in an attempt to identify those distributions that gave better fits and eliminate from consideration the ones that gave poor fits to the data. This was a scoping calculation to identify which distribution could be eliminated from consideration. The cases selected for this initial study were somewhat arbitrary but were chosen to capture a representative cross-section of the radionuclides under consideration. This initial screening considered a representative cross-section of radionuclides (with the exception of ^{137}Cs and ^{90}Sr for which data were not available for use in the initial effort). Because of their importance to dose in previous TSPA-VA calculations, ^{99}Tc, ^{129}I, and ^{237}Np were included (DOE 1998). However, as the degree of soil build-up (Table 5) for each of these radionuclides was trivial, only the data for period 6 were fitted. In addition to these three, ^{229}Th and ^{243}Am were selected as these two radionuclides were demonstrated to have significant soil build-up effects (Table 5). For these two radionuclides, irrigation periods 1 and 6 were used with all the distributions as they represented extremes of the BDCFs distributions. Because five isotopes of uranium had been considered when generating the BDCF data, it was thought prudent to include one of these in this initial screening. ^{238}U (period 6) was therefore considered.

These calculation are presented in files named "*Element Symbol*" + "*Atomic Weight*" + "Fitting Mk2" + " *Date*". The calculational approach for each sheet is given below:

1. The appropriate raw BDCF data from the files noted in 6.2.1.2 were "copied" and "pasted" into column B starting at row 5.

2. The minimum value of the BDCF was determined using the "MIN" function.
3. A scaling factor (power of 10) was determined that when used to multiply the set of BDCF data gave values greater than unity. (This was done as scoping work using the Excel function SOLVER had demonstrated that solution finding could need excessive operator intervention when small numbers are generated as optimal parameters.)
4. The data were so scaled and the scaling factor noted (note that this factor has no effect on the goodness of fit test) (Column E).
5. To provide an initial estimate of parameters of the shifted lognormal distribution, the natural log of the scaled data values above the minimum value multiplied by a factor of 0.99 was determined (Column F). The constant factor, which was close to but below unity (0.99), was used to avoid an error (log of zero) being reported for the case of the minimum value. The minimum value of the scaled data was used as a starting point for the lognormal distribution translation (*s*).
6. To aid in the initial guess of parameters (required by SOLVER), the mean and standard deviation of the scaled values (cells D7 & D9) and the natural logarithms of the shifted values (cells G2 & G3) were determined using the Excel functions "AVERAGE" and "STDEV".
7. Cells in column I7 to I17 were filled with the sequence 0.0, 0.1, 0.2, 0.3, …., 1.0.
8. The Excel function "PERCENTILE" was used to generate the boundaries of the ten bins such that each bin contained 13 observations from the raw BDCF data.
9. This table of percentile values was copied into the appropriate working space for each distribution to be tested. This copying was done in the new cell by referencing the cell in the original table (i.e., cell J32 contains the expression +J12).
10. The built in Excel cumulative distribution function appropriate to the distribution being considered (for the triangular distribution see below) was used in Column J to calculate the expected cumulative distribution. (The shifted lognormal distribution is simply the lognormal distribution with all data points offset by a constant value. Therefore the shifted lognormal distribution did not required an explicit Excel function but made use of the available lognormal distribution.) This calculation used parameters values in the cells defined above each distribution in column L or (for the shifted lognormal) M. The initial parameter values inserted in these cells either were estimates from 6 above or were intuitive guesses. These cells are the cells used by the SOLVER function to find the minimum of the χ^2 goodness of fit parameter.
11. For the triangular distribution (no built in function available), the calculation was a little more protracted. Column L was used to determine whether the particular x value in use was below or above the mode (the value of "b" in Figure 1). Columns M and N were used to calculate the expressions given in 6.1.1.1 with $x > a$ and $x < c$. Depending on the value in column L, the appropriate cdf was selected and multiplied by 130 (the total number of BDCFs), to give the expected cdf.

12. The next column (column M except for the triangular distribution where, because of additional columns being used to calculate the cdf, column P is used) then takes the difference between adjacent cdf values to give the expected number of observations in that bin. The final column then generates the value of $\frac{(observed - expected)^2}{expected}$ for each bin. The Excel SUM function is used to calculate the value of χ^2 for the parameter values defining the particular distribution.

13. The Excel SOLVER function was then used to optimize the values of the parameters defining the distribution. This is done by instructing the SOLVER function to minimize the cell containing the χ^2 value by changing the values in the cell defined by the parameters of the distribution. (NOTE to users inexperienced in using the Excel function SOLVER – if the initial estimates of the parameters are far from their optimal values, the SOLVER may not be able to find the optimal solution. This effect can be circumvented by performing a first cut estimate by optimizing the solution, one parameter at a time. Once an approximate optimal solution is achieved, all parameters can then be used to generate the optimal value.)

14. This process was repeated for each of the distributions considered.

The summary of the χ^2 distribution fitting process is given in Table 6. As discussed in Section 6.1.3, the number of degrees of freedom is one less than the difference between the number of classes ($k = 10$) and the number of parameters that specify the distribution ($p = 2$ or 3 depending on the distribution).

Table 6. Values of χ^2 for the Defined Distributions with Optimized Parameters

	Distribution				
	Normal	**Lognormal**	**Shifted lognormal**	**Triangular**	**Weibull**
Degree of Freedom	**7**	**7**	**6**	**6**	**7**
^{238}U/6	10.86	6.19	4.39	6.68	20.64
^{229}Th/1	10.73	5.33	2.42	3.77	20.26
^{229}Th/6	12.10	9.00	8.78	11.33	14.39
^{99}Tc/6	36.79	16.98	7.16	17.95	42.91
^{237}Np/6	12.43	7.27	4.61	6.49	22.10
^{243}Am/1	10.42	5.08	2.02	3.29	20.13
^{243}Am/6	13.17	9.70	9.52	11.37	19.63
^{129}I/6	7.44	5.23	5.20	5.95	12.73

As the accept/reject value for the 90% confidence level for χ^2 is 12.0 for 7 degrees of freedom (dof) and 10.6 for 6 dof, only the Weibull is judged unacceptable for every case. If the observations were from the distribution then, at the 90% level, it would be expected that if the "experiment" was repeated multiple times, the defined value χ^2 of would be exceeded 10% of the time. The 95% level are 12.6 for 6 dof and 14.1 for 7 dof (see Table 1). With the exception of ^{99}Tc for which only the shifted lognormal distribution is adequate, the other four distributions are acceptable. See Section 6.1.3 for additional discussion on the confidence level and the number of observations that can be expected to "fail" the accept/reject test.

However, there are benefits to using either a normal or a lognormal distribution. Not least of these are the intuitive meaning of the mean and standard deviation of these distributions. An additional benefit of using one of these distributions will be utilized in the successor AMR that abstracts the previous periods of irrigation for inclusion into the stochastic TSPA code. Although the normal distribution would not be rejected (excepting ^{99}Tc), observation of the values of χ^2 in Table 6 shows that the lognormal distribution provides a systematically better fit to the BDCF data. The remaining analysis in this report will focus on the lognormal and shifted lognormal distributions.

6.2.1.6 Distribution Parameter Determination

The worksheets in the Excel files discussed in 6.2.1.2 that contained the individual BDCFs data were used to optimize the fit to both the lognormal and the shifted lognormal distributions. The approached followed that defined in the previous paragraph but was only performed for the two selected distributions. The Excel spreadsheet template (Distribution Fitting Template V1_0) for performing this evaluation is documented in Attachment II as required by AP-SI.1Q.

For the radionuclides (^{229}Th, ^{137}Cs, ^{90}Sr, ^{243}Am, and ^{232}U) that had shown significant soil build-up effect, each period of pervious irrigation was analyzed. ^{227}Ac only showed a soil build-up effect of about 1 percent. This radionuclide, as an example of one that exhibited little build-up, was analyzed at every time period to determine whether there were any changes in the standard deviation of the distribution. For the same reason, ^{241}Am, ^{237}Np, ^{239}Pu, and ^{240}Pu were evaluated at the first and last period of previous irrigation. The remaining radionuclides were only evaluated after irrigation period 6. As soil build-up of radionuclides can only increase the dose for a given concentration, the use of the maximum irrigation period was considered conservative. Table 3, in Section 6.2.1.3, shows that for these radionuclides the change in the mean value of the BDCFs between period 5 and period 6 is about or less than 2%. Where there was about 15% (or less) change in their BDCF mean values over the irrigation periods considered, the conservative assumption was made that the distribution appropriate for period 6 would be used for TSPA dose calculation.

The results of the calculations are given in Table 7. It should be noted that Table 7 presents parameter values to four decimal places. Statistically this precision is unjustified. However, as (some of) these data will be used in a subsequent AMR, the intent here was not to introduce rounding errors that propagate through several AMR before being used in the TSPA dose calculations.

It should be noted that the χ^2 values for lognormal distribution on ^{14}C, ^{90}Sr, ^{99}Tc, and ^{137}Cs are significantly above the acceptance limit at the 90 percent confidence level (Section 6.1.3) of 12.0. For these radionuclides, the shifted lognormal distribution provides a better fit than the lognormal distribution. In a few cases for ^{90}Sr and ^{137}Cs, the χ^2 value is above the 90 percent acceptance level. This is expected. Suppose a set of experiments were conducted, each taking a sample of 130 random observations from a shifted lognormal distribution. Then it would be expected that one tenth of the tests would have a χ^2 value above the 90% acceptance level. In the data presented in Table 7, there are 52 individual experiments. Thus, five distributions would be expected to yield a χ^2 value higher than the acceptance limit. In Table 7 there are five such observations. In a similar manner only two percent are expected to yield a χ^2 value greater than 14.4; one of the 52 tests yielded this value. For all other radionuclides for which BDCFs have been generated, the lognormal distribution is statistically acceptable.

Table 7. Scale Factors and Best-Fit Parameters for BDCF Distributions Considered

Radionuclide/ Irrigation period	Scale Factor	Distribution						
		Lognormal			Shifted lognormal			
		Log Mean	Log SD	χ^2	Shift (s)	Log Mean	Log SD	χ^2
^{227}Ac/1	1.E-02	0.5760	0.1512	5.9	1.1073	-0.4250	0.4071	2.5
^{227}Ac/2		0.5786	0.1511	6.2	1.1088	-0.4201	0.4057	2.8
^{227}Ac/3		0.5803	0.1508	6.7	1.1223	-0.4367	0.4122	3.1
^{227}Ac/4		0.5834	0.1507	5.3	1.0996	-0.3925	0.3958	2.2
^{227}Ac/5		0.5856	0.1503	5.3	1.0958	-0.3809	0.3911	2.2
^{227}Ac/6		0.5884	0.1504	5.0	1.0880	-0.3623	0.3853	2.1
^{241}Am/1	1.E-03	1.5216	0.1511	5.1	2.7981	0.5526	0.3942	2.0
^{241}Am/6		1.6118	0.1451	7.6	3.4101	0.4337	0.4631	2.2
^{243}Am/1	1.E-03	1.5196	0.1511	5.1	2.7883	0.5529	0.3935	2.0
^{243}Am/2		1.7045	0.1317	10.0	3.7771	0.5120	0.4236	5.6
^{243}Am/3		1.8546	0.1351	8.9	1.3103	1.6244	0.1695	9.0
^{243}Am/4		1.9726	0.1523	4.3	-5.3180	2.5283	0.0875	3.5
^{243}Am/5		2.0694	0.1684	5.1	-3.5972	2.4460	0.1153	4.8
^{243}Am/6		2.1529	0.1819	9.7	1.9807	1.8879	0.2375	9.5
^{14}C/6	1.E-06	1.3968	0.0579	19.6	3.4675	-0.5913	0.4172	11.7
^{137}Cs/1	1.E-05	2.1282	0.2389	23.2	4.6694	1.2712	0.5413	17.2
^{137}Cs/2		2.3652	0.2272	15.4	4.2239	1.8372	0.3778	13.5
^{137}Cs/3		2.5378	0.2143	13.3	4.1847	2.1216	0.3209	12.3
^{137}Cs/4		2.6743	0.2048	8.4	3.6464	2.3764	0.2741	8.0
^{137}Cs/5		2.8005	0.1980	7.0	2.9345	2.5997	0.2415	6.8
^{137}Cs/6		2.9118	0.2014	7.0	1.9881	2.7951	0.2261	6.9
^{129}I/6	1.E-04	1.2703	0.1718	5.2	0.4458	1.1351	0.1965	5.2
^{237}Np/1	1.E-03	1.8975	0.1500	5.0	3.6695	1.0811	0.3365	3.1
^{237}Np/6		1.9078	0.1506	7.3	3.9754	0.9942	0.3713	4.6
^{238}Pu/6	1.E-03	1.4131	0.1490	4.5	2.5117	0.4442	0.3891	1.6
^{239}Pu/1	1.E-03	1.5033	0.1511	5.1	2.7481	0.5338	0.3944	2.0
^{239}Pu/6		1.6046	0.1402	7.9	3.4556	0.3805	0.4668	2.4
^{240}Pu/1	1.E-03	1.5016	0.1511	5.1	2.7434	0.5321	0.3944	2.0
^{240}Pu/6		1.5999	0.1407	6.7	3.3851	0.4131	0.4525	1.6
^{90}Sr/1	1.E-04	0.5626	0.1847	15.9	1.1846	-0.6045	0.5944	8.3
^{90}Sr/2		0.7412	0.2797	13.7	1.2529	-0.2547	0.7046	2.7
^{90}Sr/3		0.8684	0.3487	17.0	1.3470	-0.0768	0.8227	3.0
^{90}Sr/4		0.9525	0.4053	25.6	1.4100	0.0339	0.9122	7.7
^{90}Sr/5		1.0008	0.4278	27.6	1.4954	0.0639	0.9835	7.9
^{90}Sr/6		1.0349	0.4439	26.9	1.5248	0.1143	1.0065	7.3
^{99}Tc/6	1.E-06	1.3316	0.2593	17.0	2.1631	0.4020	0.6110	7.2
^{229}Th/1	1.E-03	1.5077	0.1541	5.3	2.7008	0.5724	0.3888	2.4
^{229}Th/2		1.9951	0.1688	14.5	-8.9786	2.7954	0.0753	13.6
^{229}Th/3		2.3025	0.2195	6.0	-13.1559	3.1459	0.0938	5.0
^{229}Th/4		2.5117	0.2456	9.8	-10.8960	3.1497	0.1292	9.0
^{229}Th/5		2.6622	0.2696	6.1	-5.0636	2.9688	0.1976	5.8

Table 7. Scale Factors and Best-Fit Parameters for BDCF Distributions Considered (Continued)

Radionuclide/ Irrigation period	Scale Factor	Distribution						
		Lognormal			Shifted lognormal			
		Log Mean	Log SD	χ^2	Shift (s)	Log Mean	Log SD	χ^2
^{229}Th/6		2.7698	0.2802	9.0	-4.1943	3.0070	0.2204	8.8
^{232}U/1	1.E-03	0.5216	0.1486	6.6	0.7302	-0.0556	0.2634	5.8
^{232}U/2		0.5701	0.1434	6.8	1.0153	-0.3039	0.3387	4.6
^{232}U/3		0.6309	0.1410	7.3	1.2515	-0.5014	0.4255	2.6
^{232}U/4		0.6718	0.1329	9.3	1.2499	-0.3686	0.3700	6.3
^{232}U/5		0.7126	0.1372	9.3	1.0775	-0.0503	0.2923	8.1
^{232}U/6		0.7433	0.1396	10.3	0.5705	0.4241	0.1918	10.1
^{233}U/6	1.E-04	1.3475	0.1494	6.3	2.2169	0.4675	0.3561	3.8
^{234}U/6	1.E-04	1.3268	0.1499	6.4	2.1383	0.4685	0.3500	4.0
^{236}U/6	1.E-04	1.2710	0.1522	7.1	1.9678	0.4494	0.3427	5.0
^{238}U/6	1.E-04	1.2562	0.1472	6.2	1.8683	0.4795	0.3167	4.4

NOTES: SD = Standard Deviation

χ^2 is dimensionless.

The units of the lognormal distribution parameters are such that when the lognormal distribution is sampled and the resultant number is multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

The units of the shifted lognormal distribution parameters are such that when the lognormal distribution is sampled, the shift value added, and the resultant number multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

7. CONCLUSIONS

7.1 FINDINGS

Stochastic BDCF data were generated by the effort reported in CRWMS M&O 2000 for each radionuclide considered of potential importance to TSPA-SR. Data were generated for six periods of prior irrigation. As a first step of incorporating the BDCF information into the TSPA code, these BDCF data have been analyzed in this AMR. The findings of this effort are listed below:

1. With the exceptions of ^{229}Th, ^{137}Cs, ^{90}Sr, ^{243}Am, and ^{232}U, the effect of radionuclide build-up in soils from prolonged irrigation has a 10% or less effect (Table 5) on mean value of the BDCFs. For these radionuclides, the BDCFs appropriate to the maximum previous period of irrigation will be used for dose calculations in TSPA.

2. For ^{229}Th, ^{137}Cs, ^{90}Sr, ^{243}Am, and ^{232}U, the effect of soil build-up and the abstraction of soil build-up will be addressed in the subsequent AMR.

3. Five statistical distributions were considered as candidates to provide abstractions for the BDCFs. The distributions considered were normal, lognormal, shifted lognormal, triangular, and Weibull. The goodness-of-fit was measured by a standard statistical

technique (χ^2). For a limited sample of BDCF data, only the Weibull distribution was unsatisfactory for the majority of radionuclides considered.

4. Of the acceptable distributions (normal, lognormal, shifted lognormal, triangular), the χ^2 values indicated that order of goodness-of-fit was shifted lognormal, lognormal, triangular, and normal. The remaining data were analyzed using the shifted lognormal and the lognormal distributions.

5. For ^{14}C, ^{90}Sr, and ^{99}Tc BDCF data, the shifted lognormal distribution provided the best fit.

6. With the exception of ^{14}C, ^{90}Sr, and ^{99}Tc, it is proposed that the BDCF abstraction be represented by a lognormal distribution.

7. For those radionuclides where build-up is to be treated conservatively, the distributions as identified in items 5 and 6 used with the optimized fitting data given in Table 8, will permit sampling the BDCFs. The sampled BDCF will reflect the uncertainty as indicated by the GENII-S code. Some implementations of the lognormal distribution require the input parameters of the mean and standard deviation to be in actual space and not *ln* space. This is achieved by raising *e* to the power of the appropriate parameter. This transformed data is given in Table 9.

8. For ^{229}Th, ^{137}Cs, ^{243}Am, and ^{232}U, where build-up effects are greater than 15 percent and the lognormal distribution is appropriate, the data to be used in abstracting soil build-up for TSPA-SR are given in Table 10. For ^{90}Sr, the appropriate distribution is the shifted lognormal with parameters as given in Table 11.

Table 8. Recommended Distributions and Parameters for Those Radionuclides That Show a Small Degree (<15%) of Soil Build-up Effects

Radionuclide/ Irrigation period	Scale Factor	Distribution				
		Lognormal		Shifted lognormal		
		Log Mean	Log SD	Shift (s)	Log Mean	Log SD
^{227}Ac	1.E-02	0.5884	0.1504			
^{241}Am	1.E-03	1.6118	0.1451			
^{14}C	1.E-06			3.4675	-0.5913	0.4172
^{129}I	1.E-04	1.2703	0.1718			
^{237}Np	1.E-03	1.9078	0.1506			
^{238}Pu	1.E-03	1.4131	0.1490			
^{239}Pu	1.E-03	1.6046	0.1402			
^{240}Pu	1.E-03	1.5999	0.1407			
^{99}Tc	1.E-06			2.1631	0.4020	0.6110
^{233}U	1.E-04	1.3475	0.1494			
^{234}U	1.E-04	1.3268	0.1499			
^{236}U	1.E-04	1.2710	0.1522			
^{238}U	1.E-04	1.2562	0.1472			

NOTES: SD = Standard Deviation

χ^2 is dimensionless.

The units of the lognormal distribution parameters are such that when the lognormal distribution is sampled and the resultant number is multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

The units of the shifted lognormal distribution parameters are such that when the lognormal distribution is sampled, the shift value added, and the resultant number multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

Table 9. Recommended Distributions and Geometric Parameters for Those Radionuclides That Show a Small Degree (<15%) of Soil Build-up Effects

Radionuclide	Scale Factor	Distribution				
		Lognormal		Shifted lognormal		
		Geometric Mean	Geometric SD	Shift (*s*)	Geometric Mean	Geometric SD
^{227}Ac	1.00E-02	1.8011	1.1623			
^{241}Am	1.00E-03	5.0118	1.1562			
^{14}C	1.00E-06			3.4675	0.5536	1.5177
^{129}I	1.00E-04	3.5619	1.1874			
^{237}Np	1.00E-03	6.7382	1.1625			
^{238}Pu	1.00E-03	4.1087	1.1607			
^{239}Pu	1.00E-03	4.9759	1.1505			
^{240}Pu	1.00E-03	4.9525	1.1511			
^{99}Tc	1.00E-06			2.1631	1.4948	1.8423
^{233}U	1.00E-04	3.8478	1.1611			
^{234}U	1.00E-04	3.7690	1.1617			
^{236}U	1.00E-04	3.5644	1.1644			
^{238}U	1.00E-04	3.5121	1.1586			

DTN:MO0003SPAABS08.004

NOTES: SD = Standard Deviation

χ^2 is dimensionless.

The units of the lognormal distribution parameters are such that when the lognormal distribution is sampled and the resultant number is multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

The units of the shifted lognormal distribution parameters are such that when the lognormal distribution is sampled, the shift value added, and the resultant number multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

7.2 TSPA-SR RECOMMENDATIONS

For those radionuclides where soil build-up effects are small for the scenario considered, the conservative BDCF distributions and parameters are presented in Table 9 (DTN:MO0003SPAABS08.004). For the remaining radionuclides that exhibit significant soil build-up effects (i.e., ^{229}Th, ^{137}Cs, ^{90}Sr, ^{243}Am, and ^{232}U), the data to be used for further abstraction analysis are given in Tables 10 and 11 (DTN:MO0003SPASEA08.005).

Table 10. Recommended Lognormal Parameters at each Period of Previous Irrigation for Those Radionuclides That Show Significant (>15%) Soil Build-up Effects

Radionuclide/ Irrigation period	**Scale Factor**	**Lognormal**	
		Log Mean	**Log SD**
^{243}Am/1	1.E-03	1.5196	0.1511
^{243}Am/2		1.7045	0.1317
^{243}Am/3		1.8546	0.1351
^{243}Am/4		1.9726	0.1523
^{243}Am/5		2.0694	0.1684
^{243}Am/6		2.1529	0.1819
^{137}Cs/1	1.E-05	2.1282	0.2389
^{137}Cs/2		2.3652	0.2272
^{137}Cs/3		2.5378	0.2143
^{137}Cs/4		2.6743	0.2048
^{137}Cs/5		2.8005	0.1980
^{137}Cs/6		2.9118	0.2014
^{229}Th/1	1.E-03	1.5077	0.1541
^{229}Th/2		1.9951	0.1688
^{229}Th/3		2.3025	0.2195
^{229}Th/4		2.5117	0.2456
^{229}Th/5		2.6622	0.2696
^{229}Th/6		2.7698	0.2802
^{232}U/1	1.E-03	0.5216	0.1486
^{232}U/2		0.5701	0.1434
^{232}U/3		0.6309	0.1410
^{232}U/4		0.6718	0.1329
^{232}U/5		0.7126	0.1372
^{232}U/6		0.7433	0.1396

DTN:MO0003SPASEA08.005

NOTES: SD = Standard Deviation

χ^2 is dimensionless.

The units of the lognormal distribution parameters are such that when the lognormal distribution is sampled and the resultant number is multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

Table 11. Recommended Shifted-Lognormal Parameters at each Period of Previous Irrigation for ^{90}Sr That Shows Significant (>15%) Soil Build-up Effects

Radionuclide/ Irrigation period	Scale Factor	Distribution		
		Shifted lognormal		
		Shift (s)	Log Mean	Log SD
^{90}Sr/1	1.E-04	1.1846	-0.6045	0.5944
^{90}Sr/2		1.2529	-0.2547	0.7046
^{90}Sr/3		1.3470	-0.0768	0.8227
^{90}Sr/4		1.4100	0.0339	0.9122
^{90}Sr/5		1.4954	0.0639	0.9835
^{90}Sr/6		1.5248	0.1143	1.0065

DTN:MO0003SPASEA08.005

NOTES: SD = Standard Deviation

χ^2 is dimensionless.

The units of the shifted lognormal distribution parameters are such that when the lognormal distribution is sampled, the shift value added and the resultant number multiplied by the scaling factor, the resultant BDCF is in units of rem/year per pico-Curie/liter.

8. INPUTS AND REFERENCES

8.1 DOCUMENTS CITED

Bulmer, M.G. 1979. *Principles of Statistics.* 3rd Edition. New York, New York: Dover. TIC: 245835.

CRWMS M&O (Civilian Radioactive Waste Management System Management and Operating Contractor)1999a. *Abstraction of BDCF Distributions (Rev. 01), ID: F2000, Activity: SSPMR280.* Work Direction and Planning Document. Las Vegas, Nevada: CRWMS M&O. ACC: MOL.19990707.0094.

CRWMS M&O 1999b. *Conduct of Performance Assessment.* Activity Evaluation, September 30, 1999. Las Vegas, Nevada: CRWMS M&O. ACC: MOL.19991028.0092.

CRWMS M&O 2000. *Non-Disruptive Event Biosphere Dose Conversion Factors.* TDP-MGR-MD-000010 REV 00. Las Vegas, Nevada: CRWMS M&O. Submit to RPC URN-0068.

DOE (U.S. Department of Energy) 1998. *Total System Performance Assessment.* Volume 3 of *Viability Assessment of a Repository at Yucca Mountain.* DOE/RW-0508. Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.19981007.0030.

DOE 2000. *Quality Assurance Requirements and Description.* DOE/RW-0333P, Rev. 9. Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.19991028.0012.

Dyer, J.R. 1999. "Revised Interim Guidance Pending Issuance of New U.S. Nuclear Regulatory Commission (NRC) Regulations (Revision 01, July 22, 1999), for Yucca Mountain, Nevada." Letter from Dr. J.R. Dyer (DOE/YMSCO) to Dr. D.R. Wilkins (CRWMS M&O), September 3, 1999, OL&RC:SB-1714, with enclosure, "Interim Guidance Pending Issuance of New NRC Regulations for Yucca Mountain (Revision 01)." ACC: MOL.19990910.0079.

Leigh, C.D.; Thompson, B.M.; Campbell, J.E.; Longsine, D.E.; Kennedy, R.A.; and Napier, B.A. 1993. *User's Guide for GENII-S: A Code for Statistical and Deterministic Simulations of Radiation Doses to Humans from Radionuclides in the Environment.* SAND91-0561. Albuquerque, New Mexico: Sandia National Laboratories. TIC: 231133.

NRC (U.S. Nuclear Regulatory Commission) 1998. *Issue Resolution Status Report Key Technical Issue: Total System Performance Assessment and Integration.* Rev. 1. Washington, D.C.: U.S. Nuclear Regulatory Commission. ACC: MOL.19990105.0083.

8.2 CODES, STANDARDS, REGULATIONS, AND PROCEDURES

64 FR(Federal Register) 8640. Disposal of High-Level Radioactive Wastes in a Proposed Geologic Repository at Yucca Mountain, Nevada. Proposed rule 10 CFR 63. Readily Available.

AP-2.16Q, Rev. 0, ICN 0. *Activity Evaluation.* Washington, D.C.: DOE, OCRWM. ACC: MOL.20000207.0716

AP-3.10Q, Rev.2, ICN 0. *Analyses and Models.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.20000217.0246.

AP-3.15Q, Rev. 1, ICN 1. *Managing Technical Product Inputs.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.20000218.0069.

AP-SI.1Q, Rev. 2, ICN 4. *Software Management.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.20000223.0508.

AP-SV.1Q, Rev. 0, ICN 0. *Control of the Electronic Management of Data.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.20000329.1181.

NLP-2-0, Rev. 5. *Determination of Importance Evaluations.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.19981116.0120.

QAP-2-0, Rev. 5. *Conduct of Activities.* Las Vegas, Nevada: CRWMS M&O. ACC: MOL.19980826.0209.

QAP-2-3, Rev. 10. *Classification of Permanent Items.* Washington, D.C.: U.S. Department of Energy, Office of Civilian Radioactive Waste Management. ACC: MOL.19990316.0006.

8.3 DATA, LISTED BY DATA TRACKING NUMBER

8.3.1 Input Data

MO0004SPABDCFS.001. Preliminary Biosphere Dose Conversion Factors (BDCFS) to be Used in the TSPA for SR. Submittal date: 04/10/2000. Submit to RPC URN-0227

8.3.2 Output Data

MO0003SPAABS08.004. Abstracted BDCF Distributions for Use in TSPA-SR. Submittal date: 03/21/2000.

MO0003SPASEA08.005. Abstracted BDCF Distributions for Use in Soil Erosion Analysis. Submittal date: 03/21/2000. Submit to RPC URN-0226

9. ATTACHMENTS

Attachment	**Title**
I	List of Acronyms
II	Spreadsheet Calculation To Optimize The Goodness Of Fit Of The Lognormal And Shifted Lognormal Distributions To The Stochastic BDCF Data.
III	CD Titled *Distribution Fitting to the Stochastic BDCF Data.* This CD contains files that were developed and used to support the findings in this AMR. The files are Excel spreadsheets for use in determining the abstractions for the stochastic BDCF data. The files were created on various dates from 17 December 1999 to 05 April 2000. The originator was A. J. Smith of DE&S from Performance Assessment Operations. Details of the files are given below

Directory	File Name (unique designator)	QA Designator
Dist Testing These files are used to test various distributions for acceptability.	U238 DT.xls	QA:NA
	Th229 DT.xls	QA:NA
	Tc99 DT.xls	QA:NA
	Np237 DT.xls	QA:NA
	Am243 DT.xls	QA:NA
	I129DT.xls	QA:NA
Param Det The files are used to determine the optimum parameters for the distributions used in the analysis.	Ac227 PD.xls	QA:QA
	C14 PD.xls	QA:QA
	I128 PD.xls	QA:QA
	Pu238 PD.xls	QA:QA
	Tc99 PD.xls	QA:QA
	U236 PD.xls	QA:QA
	U234 PD.xls	QA:QA
	U233 PD.xls	QA:QA
	U 238 PD.xls	QA:QA
	Am241 PD.xls	QA:QA
	Pu240 PD.xls	QA:QA
	Pu239 PD.xls	QA:QA
	U232 PD.xls	QA:QA
	Am243 PD.xls	QA:QA
	Th229 PD.xls	QA:QA
	Np237 PD.xls	QA:QA
	Sr90 PD.xls	QA:QA
	Cs127 PD.xls	QA:QA
Software Test These files provide the details of the template and the testing thereof.	Distribution Fitting Template V1_0.xls	QA:QA
	Template Testing B 05Apr00.xls	QA:QA

ATTACHMENT I

ACRONYMS AND ABBREVIATIONS

Acronyms

AMR	analysis/modeling report
BDCF	biosphere dose conversion factor
CRWMS M&O	Civilian Radioactive Waste Management Systems Management and Operating Contractor
DOE	Department of Energy
dof	degrees of freedom
DTN	data tracking number
IRSR	issue resolution status report
MGR	Monitored Geologic Repository
NRC	Nuclear Regulatory Commission
PAO	Performance Assessment Operation
QA	quality assurance
SD	standard deviation
TDMS	Technical Data Management Systems
TEDE	total effective dose equivalent
TSPA&I	Total System Performance Assessment and Integration
TSPA-SR	Total System Performance Assessment- Site Recommendation

Abbreviations

cdf	cumulative distribution function
Ci	Curie
pCi	pico-Curie
pdf	probability density function

ATTACHMENT II
SPREADSHEET CALCULATION TO OPTIMIZE THE GOODNESS OF FIT OF THE LOGNORMAL AND SHIFTED LOGNORMAL DISTRIBUTIONS TO THE STOCHASTIC BDCF DATA.

Distribution Fitting Template V1_0

File name DISTRIBUTION FITTING ROUTINE V1_0.XLS Version 1.0

OVERVIEW

This Attachment documents the Excel Spreadsheet template routine used in this AMR to determine the parameters that are associated with the optimal fit for the lognormal and shifted lognormal distributions to the stochastic BDCFs data provided by CRWMS M&O 2000 under DTN MO0004SPABDCFS.001. One hundred and thirty individual stochastic BDCF points were available in this data item for each of 16 radionuclides for each of 6 defined periods of irrigation prior to when the dose calculation is performed to generate the BDCFs.

The Excel Spreadsheet template routine that was used for all of the data in the AMR and DTN was worksheet named *Distribution Fitting Routine V1_0 Version 1.0* with file name *DISTRIBUTION FITTING ROUTINE V1_0.xls*.

This template routine was developed using Microsoft Excel 97 SR-2 running on a DELL POWEREDGE 2200 (Control Number 112375) with the Microsoft Windows NT operating system.

DESCRIPTION OF THE TEMPLATE

The Excel routine developed to process the multiple files of source data was classed as a template as a single Excel file (with extension .xls) was developed that contained the function and the "instructions" could be "cut and pasted" into the multiple files containing the data. The resulting files were given file names and worksheet tab names that were selected to uniquely identify the radionuclide for which the data is processed in each file.

The template is structured into three compartments. These compartments are discussed sequentially as processed by Excel. The following tables show the equations entered in the cells of the template.

Table 1 shows the input portion of the template. The 130 BDCF data item are to be copied into cells B4-133.

The initial operation is to determine a scaling factor, the smallest integral power of ten, that is required to scale the input to a set of numbers greater than unity. This is achieved by

- In cell D1 use the Excel MIN function to find the minimum of the 130 input data points.
- The LOG (of base 10) of this number is determined in cell D2.
- This number is rounded down to the nearest integer by use of the INT function in cell D3.
- The required scaling factor is found in cell D4, by raising 10 to the negative power of the number in cell D3.
- For each of the rows containing data in column B, the data are multiplied by the scaling factor (D4) and placed in column E.

To assist later in the calculation the functions AVERAGE in cell D7 and STDEV in cell D9 are sequentially applied to the scaled data in column E to generate the mean and standard deviation of the scaled data.

Table 1. The first block of the Template where the data are scaled to provide number greater than unity.

A	B	C	D	E
1	Annual	Min	=+MIN(B4:B133)	
2	EDE	Log10(Min)	=+LOG10(D1)	
3		Round	=+INT(D2)	Scaled Data
4	0.01840085	Multiplier	=10^-D3	=+B4*$D$4
5	0.0199124			=+B5*$D$4
6	0.0161617		Mean	=+B6*$D$4
7	0.0201825		=+AVERAGE(E4:E133)	=+B7*$D$4
8	0.01798456		SD	=+B8*$D$4
9	0.02373756		=+STDEV(E4:E133)	=+B9*$D$4
10	0.01869966			=+B10*$D$4
11	0.01642398			=+B11*$D$4
12	0.02120556			=+B12*$D$4
13	0.01622767			=+B13*$D$4
14	0.01611292			=+B14*$D$4
15	0.01603101			=+B15*$D$4
16	0.01875957			=+B16*$D$4
17	0.02545691			=+B17*$D$4
18	0.01817036			=+B18*$D$4
19	0.01767958			=+B19*$D$4
20	0.01659659			=+B20*$D$4
21	0.02243769			=+B21*$D$4
...	....			....
...	....			....
132	0.01956141			=+B132*$D$4
133	0.01217972			=+B133*$D$4

The second part if the calculation is to determine the bin boundaries such that each of 10 bins contains 13 data points. This is achieved in cells G1: H18 as shown in Table 2.

- Cell H4 uses the COUNT function to determine the number of data points considered (for the work reported here the number was 130, but to afford some flexibility for future Revisions a more general approach was adopted).
- The total number of data points from cell H4 is divided by 10 in cell H2 to give the number of data points in each bin.
- Cells G8:G18 are filled with the required percentile points starting at zero and increasing linearly by a constant 0.1.
- Cells H8:H18 are then filled with the values of the data corresponding to percentile points defined in column G. Cell H8 can be set at zero as the BDCF data by definition cannot be negative. Cells H9:H18 used the function PERCENTILE with parameters specifying the data range E4:E133 and the appropriate percentile points defined in column G.

- For use on the fitting of the second distribution, the information in block G8:H18 is copied by relative cell reference to block G27:H37 (i.e., cell G27 contain the expression "=+*G8*").

The final portion of the template is also shown in Table 2. This is where the results are generated and presented.

Cell J1 contains the mathematical expression to generate the inverse of the scaling factor derived in cell D4. This is the multiplier that converts the distribution outputs to the correct units (i.e., rem/year per pico-Curie/liter).

From the discussion on distribution screening in Section 6.2.1.5, the two distributions to be considered for representing the data were the lognormal and the shifted lognormal. In Table 2, the block of cells I4:K20 were used for the optimization of parameters for the shifted lognormal distribution. The lognormal distribution fitting was done in cells J24:L39. Because of the similarity of these two blocks, only one, the shifted lognormal distribution fitting, will be presented here. The description of the lognormal calculation is the same with the exception of the shift being zero and as such is not allocated its own cell.

An initial estimate of the shift of the lognormal distribution is put into cell J3.
An initial estimate of the mean of the lognormal distribution is put into cell J4.
An initial estimate of the standard deviation of the lognormal distribution is put into cell J5.

Cells J8:J18 contain the expected cumulative distribution of the 130 samples for the shifted lognormal distribution with the parameters defined in cells J3:J5 and boundaries as defined in cells H8:H18. This is performed as follows.

- Cell J8 is loaded with zero (there are no BDCFs below zero).
- Cells J9:J18 contain the product of cell H1 (the number of BDCFs samples, 130 in this case) and function LOGNORMALDIST of parameters (a) the bin edge (from the corresponding cell in range H9:H18) less the defined shift (cell J3), (b) the mean value as defined in cell J4, and (c) the standard deviation as defined in cell J5.
- Cells K9:K18 calculate the number of samples expected in the bins defined the values in cells H8:H18. As cells J8:J18 contain the expected cumulative distribution for the defined parameters, the expected value of observations in cell K9 is calculated from the expression "+*J9-J8*". This expression is copied into cells K10:K18.

Cells L9:L18 contain the expression to evaluate the square of the difference between the expected and observed number of BDCFs in the bin divided by the expected value. In cell L9, the expression is "=+*(K9-$H$2)^2/K9*". (The contents of H2 is 13.) This expression is copied into cells L10:L18.
The SUM function is used in cell L20 to form the sum of the values of cells L9:L18. The resulting value is the parameter χ^2 used to determine goodness of fit.

USE OF THE TEMPLATE

To determine the optimum parameter set for the lognormal and shifted lognormal distributions the following operations were performed on each individual set of 130 BDCFs realizations. (For logistic reasons the six data sets of BDCFs corresponding to the six period of previous irrigation for each radionuclide were loaded into six separate worksheets of a spreadsheet. The Excel file was given a self-explanatory name. The tab on each sheet was given a unique name.)

Data – Template Integration

The BDCF data were loaded into a otherwise blank sheet in cells B4:B133.
The template file (cells C1:L133) was copied to the clip board.
The cursor was placed on cell C1 of the spreadsheet containing the data.
The contents of the clip board were pasted into the target cell.

Initial Action

The initial estimates of the parameters were loaded into cells J3:J5 for the shifted lognormal distribution and J24:J25 for the lognormal. It should be noted that these estimates need only be approximate. Reasonable estimates could be obtained as follows

- for shift in the shifted lognormal, the minimum of the scaled data,
- for the mean of the lognormal, the natural log of the upper limit of the third bin,
- for the mean of the shifted lognormal, the natural log of the difference between the upper limit of the third bin and the minimum value,
- for the standard deviation of the lognormal, the natural log of the ratio between the upper limits of third and sixth bin limits,
- for the standard deviation of the shifted lognormal, the natural log of the ratio between the upper limit of third bin less the minimum value and the upper limit of the sixth bin less the minimum value.

Optimization

The Excel "Solver" was initiated. This is found on the pull down menu for "Tools"
In the "Solver Parameters" box, the following actions are taken.

1. "Set Target Cell" to L20 for the shifted lognormal distribution (L39 for the lognormal).
2. Set the "Equal to" radio button to "Min".
3. Set the "By Changing Cells" to K3:K5 for the shifted lognormal (J24:J25 for the lognormal).
4. Push the "Options" button and check the "Use Automatic Scaling" box, push "OK" button.
5. Push the "Solve" button.

If the "Solver" finds a minimum, accept the values.
If the "Solver" is unable to converge to a solution, reset the values to their initial values.
Repeat the sequence 1 to 5 above but in item 3, select only one of the parameters to vary. Repeat the last step but using a different parameter to vary. Once solutions have been reached by varying each parameter in turn, vary two at a time. Finally use the initial approach and generate a solution by varying all three parameters together.

Demonstration of functionality

To provide a demonstration that the template performs the distribution fitting to the data, two tests were conducted. The first test focused only on the distribution fitting part of the template while the second exercised to complete template.

In the first test, the bin boundaries for each 10 percentile points (i.e., 10, 20, 30, ...) were generated using the Excel function LOGINV with defined mean and standard deviation. The values used for the mean (2.345) and standard deviation (0.246) were completely arbitrary selections to generate the data to test the template. Because of the numerical approximations used by the functions in Excel, 99.9999% was used in place of 100%, which if used, gave an error condition. These data were then used to exercise the fitting portion of the template. These data being representative of the lognormal distribution were also used to test the shifted lognormal distribution (expectation was zero shift with the mean and standard deviation the same as for the lognormal distribution). Table 3 shows the spreadsheet after the Solver routine has been exercised.

For both distributions the values of χ^2 was very low. The estimated values of the parameters for each distribution were close to the actual values used to generate the data.

For the complete test of the template, the Excel spreadsheet was used to generate 130 random samples a lognormal distribution using the same parameters as above. This was achieved by generating 130 row of uniform (over 0,1) numbers with the RAND() function. The LOGINV function using the parameters of the random number, the mean, and the standard deviation, generated the 130 samples from the required distribution. The resultant numbers were entered into a new sheet starting in cell B4. The data entry was performed using the COPY followed by PASTE SPECIAL with Values selected. These 130 random samples are given in Table 4. These data are presented here as they would be needed to duplicate the work shown here.

The calculation template was imported by use of the COPY and the PASTE instructions into the sheet containing the data. The optimizations were performed as discussed above. (It should be noted that using initial estimate of the parameters that were well removed from the actual values did require the approach of using single parameter optimization. Once the values of χ^2 was below about 100, the three parameter optimization could be used.)

After fitting optimization the spreadsheet appeared as shown in Table 5. As can be seen the use of stochastic input has caused an expected increase change in the value of χ^2 over the value found for the deterministic distribution. As expected, both distributions provide acceptable fits to the random data (Section 6.1.3). The predicted means and standard deviations values of the distribution parameters are within a few percent of the actual values used to generate the stochastic data. The shift (of 0.32) predicted for the shifted lognormal distribution should be put into perspective, as the (arithmetic) mean of the stochastic data is 10.8. This shift is thus only a few percent of the distribution mean.

Conclusion

The BDCF data to be analyzed are in their own right stochastic. If the same problem were to be run multiple times using different random number seeds, it would be expected to see fluctuations in the mean and standard distributions of the outputs. Each set of BDCF outputs would have differing degrees of goodness of fit. Given the stochastic nature of the problem, the template developed for this work and documented in this attachment, has been shown to provide a sound approach to statistical distribution selection (lognormal and shifted lognormal) and associated parameter determination.

Table 2. Details of the portion of the template where the distribution parameters are fitted to the binned data.

	G	H	I	J	K	L
1	Number	=+COUNT(E4:E133)		**=1/D4**	**Multiplier**	
2	Number per bin	=+H1/10				
3				1.10725701298249	Off-set	
4				-0.424965807314677	Mean	
5	**Shifted lognormal**			0.407137834081862	SD	
6						
7	%ile	Values		Cum	Bin	
8	0	0		0		
9	0.1	=PERCENTILE($E$4:$E$133,G9)		=$H$1*LOGNORMDIST(H9-$J$3,$J$4,$J$5)	=+J9-J8	=+(K9-$H$2)^2/K9
10	0.2	=PERCENTILE($E$4:$E$133,G10)		=$H$1*LOGNORMDIST(H10-$J$3,$J$4,$J$5)	=+J10-J9	=+(K10-$H$2)^2/K10
11	0.3	=PERCENTILE($E$4:$E$133,G11)		=$H$1*LOGNORMDIST(H11-$J$3,$J$4,$J$5)	=+J11-J10	=+(K11-$H$2)^2/K11
12	0.4	=PERCENTILE($E$4:$E$133,G12)		=$H$1*LOGNORMDIST(H12-$J$3,$J$4,$J$5)	=+J12-J11	=+(K12-$H$2)^2/K12
13	0.5	=PERCENTILE($E$4:$E$133,G13)		=$H$1*LOGNORMDIST(H13-$J$3,$J$4,$J$5)	=+J13-J12	=+(K13-$H$2)^2/K13
14	0.6	=PERCENTILE($E$4:$E$133,G14)		=$H$1*LOGNORMDIST(H14-$J$3,$J$4,$J$5)	=+J14-J13	=+(K14-$H$2)^2/K14
15	0.7	=PERCENTILE($E$4:$E$133,G15)		=$H$1*LOGNORMDIST(H15-$J$3,$J$4,$J$5)	=+J15-J14	=+(K15-$H$2)^2/K15
16	0.8	=PERCENTILE($E$4:$E$133,G16)		=$H$1*LOGNORMDIST(H16-$J$3,$J$4,$J$5)	=+J16-J15	=+(K16-$H$2)^2/K16
17	0.9	=PERCENTILE($E$4:$E$133,G17)		=$H$1*LOGNORMDIST(H17-$J$3,$J$4,$J$5)	=+J17-J16	=+(K17-$H$2)^2/K17
18	1	=PERCENTILE($E$4:$E$133,G18)		=$H$1*LOGNORMDIST(H18-$J$3,$J$4,$J$5)	=+J18-J17	=+(K18-$H$2)^2/K18
19						
20					Chi Squared	=SUM(L9:L18)
21						
22						
23						
24	**lognormal**			0.57595959386264	Mean	
25				0.151219786027288	SD	
26	%ile	Values				
27	=+G8	=+H8				
28	=+G9	=+H9		=$H$1*LOGNORMDIST(H28,$J$24,$J$25)	=+J28-J27	=+(K28-$H$2)^2/K28
29	=+G10	=+H10		=$H$1*LOGNORMDIST(H29,$J$24,$J$25)	=+J29-J28	=+(K29-$H$2)^2/K29
30	=+G11	=+H11		=$H$1*LOGNORMDIST(H30,$J$24,$J$25)	=+J30-J29	=+(K30-$H$2)^2/K30
31	=+G12	=+H12		=$H$1*LOGNORMDIST(H31,$J$24,$J$25)	=+J31-J30	=+(K31-$H$2)^2/K31
32	=+G13	=+H13		=$H$1*LOGNORMDIST(H32,$J$24,$J$25)	=+J32-J31	=+(K32-$H$2)^2/K32
33	=+G14	=+H14		=$H$1*LOGNORMDIST(H33,$J$24,$J$25)	=+J33-J32	=+(K33-$H$2)^2/K33
34	=+G15	=+H15		=$H$1*LOGNORMDIST(H34,$J$24,$J$25)	=+J34-J33	=+(K34-$H$2)^2/K34

	G	H	I	J	K	L
35	=+G16	=+H16		=$H$1*LOGNORMDIST(H35,$J$24,$J$25)	=+J35-J34	=+(K35-$H$2)^2/K35
36	=+G17	=+H17		=$H$1*LOGNORMDIST(H36,$J$24,$J$25)	=+J36-J35	=+(K36-$H$2)^2/K36
37	=+G18	=+H18		=$H$1*LOGNORMDIST(H37,$J$24,$J$25)	=+J37-J36	=+(K37-$H$2)^2/K37
38						
39					Chi Squared	=SUM(L28:L37)

Table 3. The spreadsheet after parametric optimization on the deterministic distribution data.

Number	130			
Number per bin	13	Optimized Parameters		
Mean	**2.345**	**0.003158**	Off-set	
SD	**0.246**	**2.344691**	Mean	
Shifted lognormal		**0.246087**	SD	
%ile	Values	Cumulative Number	Number in Bin	
0	0	0.0000		
0.1	7.612	13.0005	13.0005	0.0000
0.2	8.482	26.0014	13.0009	0.0000
0.3	9.171	39.0018	13.0004	0.0000
0.4	9.803	52.0018	13.0000	0.0000
0.5	10.433	65.0013	12.9995	0.0000
0.6	11.104	78.0004	12.9992	0.0000
0.7	11.870	90.9995	12.9991	0.0000
0.8	12.833	103.9985	12.9990	0.0000
0.9	14.300	116.9978	12.9993	0.0000
0.999999	33.718	129.9999	13.0021	0.0000
			Chi Squared	**6.73E-07**
		Optimized Parameters		
Lognormal		**2.345000**	Mean	
		0.245999	SD	
%ile	Values	Cumulative Number	Number in Bin	
0	0			
0.1	7.612096	13.0000	13.0000	0.0000
0.2	8.482126	26.0000	13.0000	0.0000
0.3	9.170551	39.0000	13.0000	0.0000
0.4	9.802885	52.0001	13.0001	0.0000
0.5	10.43327	65.0001	13.0000	0.0000
0.6	11.1042	78.0001	13.0000	0.0000
0.7	11.86986	91.0002	13.0001	0.0000
0.8	12.83324	104.0001	13.0000	0.0000
0.9	14.30003	117.0001	12.9999	0.0000
0.999999	33.71758	129.9999	12.9998	0.0000
			Chi Squared	**4.74E-09**

Table 4. The stochastic lognormal data points generated for use in the template test.

Data Point	Log-normal Sample	Data Point	Log-normal Sample	Data Point	Log-normal Sample	Data Point	Log-normal Sample	Data Point	Log-normal Sample
1	8.145736	27	10.40769	53	17.71287	79	16.69246	105	9.838233
2	7.809803	28	11.92897	54	10.42232	80	7.930648	106	4.956826
3	11.92865	29	10.43261	55	11.06095	81	12.20572	107	11.06202
4	7.601629	30	12.35295	56	10.43778	82	12.84098	108	14.44208
5	8.332108	31	9.958758	57	11.50695	83	9.085715	109	9.150307
6	6.263796	32	10.13731	58	13.63304	84	11.0794	110	17.60808
7	11.61586	33	12.80746	59	13.54673	85	11.48129	111	9.684837
8	7.668111	34	13.22788	60	8.538925	86	8.046077	112	13.549
9	13.74118	35	7.80654	61	10.33902	87	10.37932	113	6.554172
10	11.52526	36	12.42003	62	7.534568	88	8.343525	114	8.31692
11	9.168666	37	9.56927	63	7.83895	89	8.212199	115	9.287142
12	16.84451	38	10.31187	64	9.990196	90	7.646379	116	19.3567
13	9.525904	39	13.38267	65	10.38994	91	7.960242	117	8.715749
14	7.481373	40	11.05763	66	12.02133	92	15.4727	118	6.448932
15	9.975527	41	7.834598	67	8.760372	93	14.94796	119	12.6583
16	7.511643	42	13.01222	68	10.45329	94	11.02352	120	12.43678
17	8.965911	43	7.229773	69	13.74181	95	8.342806	121	8.098561
18	10.23618	44	8.001024	70	10.54139	96	11.80817	122	13.815
19	12.89763	45	13.41179	71	9.441505	97	13.54648	123	12.04671
20	9.553955	46	8.451972	72	23.82781	98	11.14905	124	8.746902
21	12.52213	47	11.73317	73	10.83975	99	11.58038	125	12.2077
22	14.89621	48	9.214154	74	8.789197	100	11.78174	126	9.610761
23	12.07468	49	9.080705	75	10.72109	101	6.910238	127	12.58278
24	14.07942	50	7.80207	76	8.897759	102	7.481009	128	14.10004
25	10.83692	51	8.803638	77	11.46085	103	10.53823	129	9.656757
26	11.18526	52	13.92751	78	13.05103	104	8.708254	130	11.8252

Table 5. The spreadsheet after parameter optimization on the 130 items of stochastic data.

B	C	D	E	F	G	H	I	J	K	L
	Min	4.96E+00			Number	130		**1.0E+00**	**Multiplier**	
	Log10(Min)	6.95E-01			Number per bin	13				
	Round	0.00E+00	Scaled Data					**0.3231**	**Off-set**	
8.145736	**Multiplier**	**1.00E+00**	8.1457					**2.308296**	**Mean**	
7.809803			7.8098		**shifted lognormal**			**0.243830**	**SD**	
11.92865		Mean	11.9286							
7.601629		10.801	7.6016		%ile	Values		Cum	Bin	
8.332108		SD	8.3321		0	0		0.0000		
6.263796		2.830	6.2638		0.1	7.789		14.4071	14.4071	0.1374
11.61586			11.6159		0.2	8.329		22.7181	8.3110	2.6456
7.668111			7.6681		0.3	9.046		36.3634	13.6453	0.0305
13.74118			13.7412		0.4	9.777		51.9768	15.6134	0.4374
11.52526			11.5253		0.5	10.435		66.1554	14.1786	0.0980
9.168666			9.1687		0.6	11.164		80.7010	14.5456	0.1642
16.84451			16.8445		0.7	11.957		94.2223	13.5213	0.0201
9.525904			9.5259		0.8	12.852		106.1170	11.8947	0.1027
7.481373			7.4814		0.9	13.826		115.2498	9.1328	1.6375
9.975527			9.9755		1	23.828		129.9676	14.7178	0.2005
7.511643			7.5116							
8.965911			8.9659						**Chi Squared**	**5.473983555**
10.23618			10.2362							
12.89763			12.8976							
9.553955			9.5540							
12.52213			12.5221		**lognormal**			**2.340325**	**Mean**	
14.89621			14.8962					**0.236081**	**SD**	
12.07468			12.0747		%ile	Values				
14.07942			14.0794		0	0				
10.83692			10.8369		0.1	7.788675		14.4983	14.4983	0.1548
11.18526			11.1853		0.2	8.329071		22.7593	8.2610	2.7186
10.40769			10.4077		0.3	9.046267		36.3305	13.5712	0.0240
11.92897			11.9290		0.4	9.776874		51.8948	15.5643	0.4225

B	C	D	E	F	G	H	I	J	K	L
10.43261			10.4326		0.5	10.4352		66.0674	14.1727	0.0970
12.36295			12.3530		0.6	11.16354		80.6439	14.5765	0.1705
9.958758			9.9588		0.7	11.95668		94.2216	13.5777	0.0246
10.13731			10.1373		0.8	12.85231		106.1783	11.9567	0.0910
12.80746			12.8075		0.9	13.82625		115.3537	9.1753	1.5943
13.22788			13.2279		1	23.82781		129.9676	14.6139	0.1782
7.80654			7.8065							
12.42003			12.4200						**Chi Squared**	**5.475564356**
9.56927			9.5693							
10.31187			10.3119							
13.38267			13.3827							
11.05763			11.0576							
7.834598			7.8346							